Vers une construction collaborative de bases d'amers géo-référencées pour la localisation en ligne d'un véhicule en milieu urbain

Dorra Larnaout, Vincent Gay-Bellile, Steve Bourgeois, Michel Dhome

Vers une construction collaborative de bases d'amers géo-référencées pour la localisation en ligne d'un véhicule en milieu urbain

D. Larnaout[1] V. Gay-Bellile[1] S. Bourgeois[1] M. Dhome[2]
[1] CEA, LIST, LVIC, Point Courrier 173, F-91191, Gif-Sur-Yvette, France
[2] Institut Pascal, UMR 6602 Université Blaise Pascal/CNRS/IFMA
[1] prénom.nom@cea.fr, [2] michel.dhome@univ-bpclermont.fr

Résumé

L'aide à la navigation par Réalité Augmentée RA nécessite une estimation précise des six paramètres de la camera. Pour ceci, les solutions de localisation par vision passent par une étape de modélisation hors ligne de l'environnement. Tandis que les solutions existantes exigent des matériels coûteux et/ou un temps d'exécution très important, nous proposons dans cet article un processus qui crée automatiquement une modélisation précise de l'environnement en utilisant uniquement une caméra standard, un GPS bas coût et des modèles SIG Système d'Informations Géographique disponible gratuitement.

Mots Clef

SLAM monoculaire, GPS, modèles SIG.

Abstract

To provide high quality Augmented Reality AR service in a car navigation system, accurate 6DoF localization is required. To ensure such accuracy, most of current vision-based solutions rely on an off-line large scale modeling of the environment. Nevertheless, while existing solutions require expensive equipments and/or a prohibitive computation time, we propose in this system paper a complete framework that automatically builds an accurate city scale database of landmarks using only a standard camera, a GPS and GIS Geographic Information System.

Keywords

Monocular SLAM, GPS, GIS models.

1 Introduction

Afin d'offrir des services d'aide à la navigation par RA, des solutions de localisation basées vision peuvent être utilisées [5, 6, 3]. Ces solutions nécessitent souvent la création préalable d'une base précise d'amers géo-référencés à grande échelle. Cette base peut être par la suite utilisée pour assurer une localisation en ligne. Pour créer la base en question, ces solutions exploitent généralement soit des matériels couteux (LIDAR, GPS RTK,...) soit des collections d'images disponibles sur internet, souvent concentrées sur des sites d'intérêt (tour Eiffel, le Colisée, ...). Or, pour qu'une base d'amers soit exploitable dans le cadre de la localisation de véhicule, celle-ci doit être précise et couvrant des points de vues semblable à ceux des utilisateurs finaux [1]. De plus, elle doit être mise à jour le plus régulièrement possible afin d'intégrer les changements intervenant sur l'environnement.

Au vu de ces critères, seule une approche collaborative où les utilisateurs participent à la création de la base semble envisageable. En effet, la multiplicité des utilisateurs permet de couvrir la totalité du territoire à traiter sous l'ensemble des points de vues et conditions d'illumination souhaitées. Toutefois, une telle approche implique l'exploitation de capteurs bas coût et des calculateurs accessibles. Ainsi, seuls des capteurs largement répandus (GPS standard, caméras,...) sont tolérés. D'autre part, si une partie du calcul peut être déportée sur un serveur, la quantité d'informations à transférer à celui-ci ainsi que le temps de traitement doivent être réduits au maximum étant donné le nombre d'utilisateurs. Pour ceci, nous proposons dans cet article une solution qui exploite uniquement un GPS standard, une caméra bas coût et des modèles SIG largement disponibles. Plus précisément, la base d'amers est créée à travers la fusion du SLAM visuel avec les données GPS et le MET (Modèle d'Elévation de Terrain). Ensuite, de par son volume limité, la base résultante peut être transférée à un serveur où elle sera optimisée en exploitant les contraintes géométriques fournies par le MET et les modèles 3D des bâtiments. Cette solution étend les travaux de [7] et nos précédents travaux [10].

Dans ce qui suit, un état de l'art des méthodes existantes est présenté dans la section 2. Après un bref rappel de la méthode introduite dans [7], le synoptique de notre nouvelle approche est exposé dans la section 3. Dans la section 4, la solution proposée est détaillée. Enfin, des évaluations sur des données réelles ainsi qu'un exemple d'application de RA sont présentés dans la section 5.

Notation Dans cet article, deux différents repères sont utilisés : le repère monde correspondant au référentiel du GPS et le repère de plan de la route associé à chaque plan du MET *Modèle d'Élévation de Terrain* [2] où l'axe des X et l'axe des Y définissent le plan du sol tandis que l'axe

1. Point de vue d'un véhicule se déplaçant sur la route
2. Le MET utilisé est un modèle géométrique où chaque route est représentée par un plan

des Z représente sa normale. Par exemple, si $\mathbf{q_1}$ est le vecteur contenant les coordonnées du point 3D Q_1 dans le repère monde, nous notons $\hat{\mathbf{q}}_\mathbf{1}$ le vecteur contenant ses coordonnées dans le repère de plan de la route. Dans la suite, nous supposons que la reconstruction de la scène observée est estimée par le SLAM monoculaire introduit dans [2]. Il s'agit d'un SLAM basé sur le principe d'image clé et utilise un ajustement de faisceaux local pour raffiner l'erreur de re-projection de M points 3D observés dans N images clés. La fonction de coût résultante est donnée par :

$$E_1(\mathbf{x}) = \sum_{i=1}^{i=M} \sum_{j \in \mathscr{A}_i} d^2\left(u_{i,j}, KP_j\mathbf{q_i}\right), \qquad (1)$$

où K est la matrice des paramètres intrinsèques de la caméra et P_j est sa j^{me} pose. $u_{i,j}$ est l'observation 2D du i^{me} point 3D Q_i dans la j^{me} image clé. $\mathscr{A}_i$ est l'ensemble des indexes des images clés observant Q_i. $\mathbf{x}$ représente le vecteur contenant les paramètres à optimiser. Ce vecteur est organisé tel que $\mathbf{x}^T = \left(\mathbf{c_1}^T, \mathbf{c_2}^T, \mathbf{q}^T\right)$ où $\mathbf{c_1} = \left(\mathbf{t_x}^T, \mathbf{t_y}^T, \psi^T\right)$ concatène les paramètres *dans le plan* de la caméra (*i.e* le déplacement 2D et l'angle lacet), $\mathbf{c_2} = \left(\mathbf{t_z}^T, \alpha^T, \gamma^T\right)$ concatène les paramètres *hors plan* (*i.e* l'altitude de la camera, les angles roulis et tangage) tandis que $\mathbf{q}$ contient les positions 3D de tous les points observés.
Finalement, les mesures *dans le plan* fournies par le GPS sont respectivement stockées dans les vecteurs $\mathbf{t}_x^{gps}$ et $\mathbf{t}_y^{gps}$.

2 Travaux connexes

Plusieurs méthodes ont été proposées afin de créer des bases d'amers géo-référencés pour des grands environnements. Ces bases contiennent souvent un ensemble d'amers 3D que nous désignons par le terme "nuage de points" obtenu généralement par un algorithme de type *"Structure from Motion"* (SfM). Dans ce qui suit nous intéresserons en particulier aux méthodes qui géo-référencent un nuage de points 3D.
Pour assurer le géo-référencement de la base d'amers, certaines méthodes exploitent conjointement l'information du GPS disponible et la géométrie multi-vues. Certaines approches telles que celles proposées dans [3] utilisent une collection d'images géo-référencées disponibles sur internet. Toutefois, en plus du temps d'exécution très important nécessaire pour la création de la base, la précision de la géo-localisation obtenue reste limitée à celle du GPS. Pour améliorer cette précision, d'autres méthodes proposent d'exploiter en plus les informations géométriques et géographiques apportées par des modèles SIG. Ceci est réalisé en estimant des transformations rigides qui recalent au mieux la reconstruction 3D avec les empreintes de bâtiments obtenues soit à partir d'images satellite dans [5] ou directement à partir d'une carte 2D des bâtiments [6]. Toutefois, une transformation rigide est incapable de modéliser des déformations complexes. Ainsi, cette approche n'est pas suffisamment précise pour corriger les erreurs de la reconstruction SfM. D'autre part, les amers 3D obtenus à partir d'une collection d'images sur internet sont souvent focalisés sur des lieux d'intérêt observés depuis le point de vue des piétons alors qu'ils doivent être distribués uniformément sur l'ensemble du réseau routier et du point de vue du conducteur afin d'assurer une localisation précise d'un véhicule. Pour faire face à ces deux limitations, [7] propose d'une part d'exploiter une reconstruction SLAM obtenue à partir d'un flux vidéo enregistré par une caméra embarquée dans un véhicule. D'autre part, des transformations non rigides sont estimées afin de contraindre la reconstruction en question avec les modèles 3D des bâtiments. Cette solution semble mieux répondre à notre problématique. Pour cette raison, nous adoptons le même concept et nous l'étendons afin de traiter ses limitations qui seront détaillées ci-dessous.

3 Positionnement et synoptique

Dans [7], Lothe et. al ont démontré la possibilité de corriger et géo-référencer une reconstruction SLAM en exploitant des modèles 3D des bâtiments grossiers à travers des transformations non-rigides. Le principe de l'approche est d'aligner le nuage de points correspondant aux façades avec les plans 3D des modèles des bâtiments. Pour ceci, une similitude par morceaux utilisant les données GPS au niveau des virages est appliquée, dans un premier temps, afin de corriger et géo-référencer grossièrement les dérives du SLAM. Une ICP non-rigide est, par la suite, réalisée afin d'aligner d'une façon plus précise le nuage de points de la reconstruction sur les modèles 3D des bâtiments. Enfin, la base résultante (l'ensemble des poses de la caméra et le nuage de points 3D) est raffinée à travers un ajustement de faisceaux global contraint aux modèles des bâtiments. Notons que dans cette approche, seuls les points 3D représentant les façades des bâtiments sont utilisés au cours de l'ICP et l'ajustement de faisceaux. Malgré les résultats prometteurs, cette solution présente certaines limitations. En effet, son étape d'initialisation sous-exploite les données GPS et ne traite pas les variations d'altitude ce qui entraine une reconstruction initiale peu précise pouvant perturber l'étape de l'optimisation (l'ICP et l'ajustement de faisceaux global). D'autre part, utiliser uniquement des points associés aux modèles des bâtiments (*ie* qui représentent les façades) rend cette solution limitée aux zones urbaines denses. Par ailleurs, durant l'optimisation, tous les degrés de liberté sont raffinés alors que les bâtiments contraignent principalement les degrés de liberté *dans le plan*. Par conséquent, les paramètres *hors plan* (principalement l'altitude et l'angle tangage) peuvent être détériorés notamment quand la qualité de l'initialisation est mauvaise.
Pour résoudre ces problèmes, plusieurs améliorations seront introduites dans la suite de cet article. Premièrement, nous proposons de tirer plus profit du GPS afin d'améliorer l'estimation des degrés de liberté *dans le plan* au cours de l'initialisation. Pour apporter plus de précision sur les degrés de liberté *hors plan*, le MET est également exploité. Par ailleurs, dans le but d'étendre la solution de [7] aux milieux péri-urbains et garantir plus de robustesse quand peu

de bâtiments sont observables, nous proposons de prendre en compte à la fois les contraintes géométriques des points 3D associés aux modèles des bâtiments et les contraintes multi-vues fournies par l'ensemble de points 3D représentant le reste de l'environnement. Étant donné qu'il n'est pas trivial de fusionner simultanément toutes ces contraintes sans perturber la convergence du processus d'optimisation, nous proposons une solution qui se focalise dans un premier temps sur une correction grossière de la dérive du SLAM avant d'optimiser plus finement la base d'amers résultante.

4 Solution proposée

Pour permettre la création automatique d'une base d'amers géo-référencée précise, notre solution comprend trois étapes que nous allons détailler ci-dessous.

4.1 Reconstruction initiale

Un SLAM monoculaire basé sur un ajustement de faisceaux local [2] fournit en ligne une représentation 3D de l'environnement. Cependant, la reconstruction obtenue n'est pas géo-référencée et souffre souvent de dérive de facteur d'échelle et d'accumulation des erreurs.

Pour faire face à ce problème, nous choisissons d'utiliser l'approche que nous avons introduit dans [10]. Elle consiste à fusionner en ligne les mesures de GPS, les données fournies par le MET et les contraintes multi-vues dans un même processus d'optimisation. Tandis que les mesures GPS contraignent la position *dans le plan* de la caméra, le MET nous apporte des informations sur son altitude δ qui est supposée fixe par rapport au plan de la route (la caméra est rigidement embarquée dans le véhicule). Afin de garantir plus de robustesse face aux données aberrantes du GPS et aux incertitudes du MET, le processus d'optimisation utilisé se base sur un ajustement de faisceaux avec une contrainte d'inégalité inspirée de la méthode introduite dans [8]. Celle-ci est réalisée en deux étapes. La première étape consiste à effectuer un ajustement de faisceaux classique où l'erreur de re-projection standard $E_1(\mathbf{x})$ est minimisée. Au cours de la deuxième étape, une seconde optimisation non linéaire est effectuée dans laquelle, la distance entre les positions *dans le plan* de la caméra et les mesures GPS ainsi que la distance entre l'altitude de la caméra estimée par le SLAM et la hauteur souhaitée déduite à partir du MET sont minimisées. Afin, de conserver une cohérence vis à vis de la géométrie multi-vues, cette seconde optimisation intègre, en plus du terme d'accroche aux données GPS et au MET, un terme de régularisation basé sur l'erreur de re-projection $E_1(\mathbf{x})$: ce terme interdit toute dégradation de l'erreur de re-projection au delà d'un seuil prédéfini e_t basé sur le résultat de la première optimisation [3]. La fonction de coût résultante est donnée par :

$$C_I(\mathbf{x}) = \frac{\omega}{e_t - E_1(\mathbf{x})} + \left\| \begin{pmatrix} \mathbf{t_x} \\ \mathbf{t_y} \\ \hat{\mathbf{t}}_\mathbf{z} \end{pmatrix} - \begin{pmatrix} \mathbf{t}_x^{gps} \\ \mathbf{t}_y^{gps} \\ \mathbf{h} \end{pmatrix} \right\|^2, \quad (2)$$

3. Une dégradation de 5% de l'erreur de re-projection initiale.

où ω est une constante positive déterminée empiriquement et $\mathbf{h} = (\underbrace{\delta \dots \delta}_{\text{N times}})^T$. $\hat{\mathbf{t}}_\mathbf{z} = \left(\hat{t}_z^1 \dots \hat{t}_z^N\right)^T$ est le vecteur concaténant les altitudes de la caméra pour N images clés optimisées dans l'ajustement de faisceaux local. Ces paramètres sont exprimés dans leurs repères plan de route correspondant. Notons que l'association caméra-plan de route est effectuée avant l'optimisation en faisant correspondre la caméra à la route la plus proche dans le MET. Pour une convergence optimale, un processus d'optimisation itératif est adopté où les associations caméra-plan de route sont remises en cause après la minimisation de la fonction de coût par l'algorithme du Levenberg Marquardt.

Cette approche permet de créer en ligne et automatiquement une reconstruction initiale géo-référencée et cohérente. Cependant, sa précision reste limitée à l'incertitude du GPS. Pour cette raison, la base obtenue est raffinée, *a posteriori*, à travers deux ajustements de faisceaux globaux contraints à des modèles SIG (*i.e* les modèles 3D des bâtiments et le MET).

4.2 Recalage 2D exploitant les modèles 3D des bâtiments

Pour traiter les imprécisions caractérisant les degrés de liberté *dans le plan*, [7] propose d'utiliser un ajustement de faisceaux global contraint aux modèles 3D des bâtiments où seuls les points 3D associés aux modèles 3D sont exploités. Ceci peut entrainer des problèmes de convergence quand peu de bâtiments sont visibles. Pour garantir plus de robustesse face à cette limitation, nous choisissons alors d'adopter plutôt la méthode introduite par [9]. Celle-ci exploite à la fois les points 3D associés aux modèles des bâtiments $Q_i \in \mathscr{M}$ et ceux appartenant au reste de l'environnement $Q_i \in \mathscr{U}$. La fonction de coût utilisée est donc composée par deux termes. Le premier terme est associé à l'ensemble de points $Q_i \in \mathscr{U}$ où deux observations $(u_{i,j}, u_{i,k})$ d'un même point Q_i sont liées par la matrice Fondamentale. Le deuxième terme correspond à l'ensemble de points $Q_i \in \mathscr{M}$ où deux observations d'un même point Q_i, supposé appartenir à une façade de bâtiment, sont liées par une Homographie. Notons également qu'au lieu d'optimiser tous les degrés de liberté de la reconstruction comme il est proposé dans [9], seuls les paramètres *dans le plan* sont raffinés afin d'éviter les problèmes de convergence.

$$\begin{aligned} E_2(\mathbf{c_1}) = &\sum_{i \in \mathscr{U}} \sum_{(j,k) \in \mathscr{A}_i}^{j \neq k} \rho\left(d_l^2\left(u_{i,j}, F_{j,k}(\mathbf{c_1})\, u_{i,k}\right), s_1\right) \\ &+ \sum_{i \in \mathscr{M}} \sum_{(j,k) \in \mathscr{A}_i}^{j \neq k} \rho\left(d^2\left(u_{i,j}, H_{j,k}(\mathbf{c_1})\, u_{i,k}\right), s_2\right), \end{aligned} \quad (3)$$

avec $d_l(u,l)$ est la distance euclidienne entre un point et une droite et $d(u_1,u_2)$ est la distance euclidienne entre deux points. $\rho(\mathbf{v},s)$ est le M-estimateur de Geman-McClure dont le seuil s est calculé en utilisant le MAD *Median Absolute Deviation* du vecteur des résidus $\mathbf{v}$. Pour une convergence optimale, un processus d'optimisation itératif est adopté également pour cette étape où les associa-

tions point-façade de bâtiment sont remises en cause après la minimisation de la fonction de coût assurée par l'algorithme du Levenberg Marquardt.

4.3 Raffinement précis exploitant les modèles 3D des bâtiments et le MET

Une fois que le recalage dans le plan est réalisé, seules des imperfections, causées par une mauvaise estimation des degrés de liberté *hors plan*, restent notables. Pour traiter ces imperfections, un deuxième ajustement de faisceaux exploitant un modèle SIG complet est réalisé. Durant cette étape, tous les degrés de liberté de la reconstruction sont optimisés et contraints simultanément aux modèles 3D des bâtiments et au MET. Tandis que les bâtiments contraignent les degrés de liberté *dans le plan*, le MET fournit des informations relatives aux degrés de liberté restant. En effet, en plus de la contrainte explicite en altitude, le MET permet, globalement, de contraindre implicitement l'angle tangage dans les lignes droites et l'angle roulis et tangage au niveau des virages. Pour introduire progressivement les contraintes du MET sans perturber le résultat du recalage *dans le plan* effectué dans l'étape précédente, un ajustement de faisceaux avec une contrainte d'inégalité (voir le principe dans la section 4.1) est utilisé. La fonction de coût associée est composée par l'erreur de re-projection prenant en compte la contrainte des bâtiments et un terme de pénalité calculé en se basant sur la contrainte MET. Cette pénalité représente la distance entre l'altitude de chaque pose de la caméra $\mathbf{t}_j^{k_j}(z)$ exprimée dans le repère plan de route associé et l'altitude δ souhaitée. Par conséquent, la fonction de coût résultante est donnée par :

$$F_I(\mathbf{x}) = \frac{\omega}{e_t - E_2(\mathbf{x})} + \|\hat{\mathbf{t}}_\mathbf{z} - \mathbf{h}\|^2 \qquad (4)$$

avec $E_2(\mathbf{x})$ est la fonction de coût introduite dans la section 4.2 mais cette fois ci, tous les degrés de liberté sont optimisés contrairement à l'étape précédente où seuls les paramètres dans le plan sont raffinés. Le processus d'optimisation itératif introduit dans les sections 4.1 et 4.2 est adopté. Ainsi, les associations point-façade de bâtiment et caméra-plan de route sont remises en cause après la minimisation de la fonction de coût assurée par l'algorithme du Levenberg Marquardt.

5 Résultats

Dans la suite, nous proposons une évaluation de notre processus. Pour ceci, plusieurs séquences réelles présentant des trajectoires de plusieurs kilomètres de long sont utilisées. Après avoir présenté les séquences utilisées dans la section 5.1, nous évaluons l'ensemble de notre processus dans la section 5.2. Enfin, une comparaison avec la méthode de [7] est établie dans la section 5.3.

5.1 Séquences de tests utilisées

Les performances de notre méthode sont vérifiées sur trois séquences réelles enregistrées dans les quartiers de Versailles et Saclay dans des conditions de conduite normale ($50Km/h$). Pour ceci, le véhicule a été équipé par un GPS standard $1Hz$ et une caméra RGB fournissant 30 images par seconde et avec un champs de vision de 90°. Les modèles SIG utilisés ont une incertitude de $2m$.

FIGURE 1 – **Concaténation de deux bases de données différentes dans le quartier de Versailles.**

Les trois séquences utilisées représentent des longs parcours de $2400m$, $1800m$ et $1200m$. Notons que même si ces séquences ne couvrent pas tout un quartier, il est possible, comme le montre la figure 1, de fusionner plusieurs bases de données d'un même quartier pour créer une base d'amers sur l'échelle d'une ville.

5.2 Évaluation de notre solution

Pour évaluer l'ensemble du processus proposé, nous utilisons dans cette section les séquences réelles décrites ci-dessus. Uniquement quelques minutes sont nécessaires pour créer la base d'amers souhaitée en utilisant un code non optimisé [4] exécuté sur un Intel(R) Xeon(R) CPU quad cores 2.4 GHz. En effet, pour la séquence de Versailles de $2400m$ de long, la base initiale, de volume 8Mo et contenant 548 vues géo-référencées et 34178 points 3D, est obtenue en ligne. Ensuite, 50 et 90 secondes sont respectivement nécessaires pour réaliser le recalage dans le plan décrit dans la section 4.2 et le raffinement sur tous les degrés de liberté introduit dans la section 4.3.

La précision atteinte sur les degrés de liberté *dans le plan* de la base résultante est évaluée à travers des vues de dessus du nuage de points reconstruit (voir figure 3). Par ailleurs, la validité de chaque étape de notre processus est mise en évidence à travers les re-projections des modèles 3D des bâtiments sur les images de la séquence après chaque étape (voir figure 2). Ces re-projections nous informent sur la précision de tous les degrés de liberté.

Comme le montre la figure 3, la fusion du SLAM avec les données GPS et MET, décrit dans la section 4.1, permet de créer automatiquement des bases de données géo-référencées (les reconstructions rouges dans la figure 3). Toutefois, ces reconstructions initiales sont objet d'importantes imprécisions au niveau des degrés de liberté *dans le plan* comme il est montré dans la figure 3 et mis en

4. Le GPU n'est pas utilisé.

(a) (b) (c)

FIGURE 2 – **Exemples de re-projection des modèles des bâtiments.** (a) Résultat après la première étape de la création de la base de données ; (b) Résultat après la seconde étape. (c) Résultat final obtenu après la troisième étape.

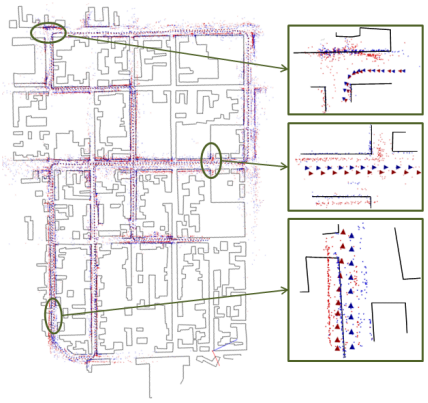


FIGURE 3 – **Validation de notre processus de création de bases d'amers géo-référencées**. Vues de dessus des bases d'amers obtenues dans le quartiers de Versailles après la première étape (SLAM contraint aux données GPS et au MET) (rouge) et après l'étape du raffinement sur tous les degrés de liberté (section 4.3) (bleu).

évidence dans la figure 2(a). Une fois le recalage *dans le plan* est effectué, ces incertitudes sont corrigées. Par conséquent, les modèles des bâtiments sont mieux alignés avec les façades observées dans le flux vidéo (voir la figure 2(b)). Enfin, le dernier raffinement exploitant le modèle SIG complet permet de corriger les imperfections restantes sur les degrés de liberté hors plan (voir figure 2(c)). Cette correction est réalisée sans perturber les paramètres dans le plan comme le montre la figure 3 où le nuage de points final représenté en bleu est parfaitement recalé sur les modèles des bâtiments.

Une fois les bases d'amers créées, elles peuvent être utilisées pour assurer une localisation en ligne. La majorité des méthodes existantes se base uniquement sur des algorithmes de reconnaissance de point de vue. Pour garantir plus de robustesse face aux changements d'illumination, nous choisissons d'adopter la méthode proposée par [1]. Cette dernière fusionne la reconnaissance de point de vue avec l'algorithme du SLAM. Comme le montre la figure 6 et la vidéo disponible en matériel supplémentaire, la précision de la localisation obtenue a permis des applications de RA.

5.3 Comparaison avec l'approche de [7]

Dans cette section, nous comparons la précision de nos bases d'amers avec celles obtenues avec la méthode de [7]. Principalement, nous souhaitons comparer les deux méthodes d'optimisation utilisées : le recalage *dans le plan* suivi par les raffinement de tous les degrés de liberté contre l'ICP non rigide suivi par l'ajustement de faisceaux global contraint uniquement aux modèles des bâtiments introduit par [7]. Pour ceci, les deux méthodes d'optimisation sont initialisées en utilisant le SLAM contraint aux données GPS et MET qui fournit une reconstruction initiale assez précise. Étant donné que la vérité terrain n'est pas disponible pour les séquences réelles utilisées, nous labellisons manuellement les coins des bâtiments dans quelques images extraites des flux vidéos (environ 20 images distribuées uniformément dans chaque séquence). Nous calculons par la suite l'erreur de re-projection entre les coins labellisés et la re-projection des coins des modèles des bâtiments. Les résultats obtenus sont résumés dans le tableau 1. La mesure d'erreur de re-projection inclue les incertitudes des modèles des bâtiments et celles de la labellisation manuelle. Cependant, ces incertitudes restent faibles par rapport à l'amélioration notable que notre solution apporte. En effet, pour les séquences de Versailles, l'approche proposée réduit de moitié la moyenne des erreurs de re-projections obtenue par l'algorithme de [7]. L'écart-type des erreurs mesurées a également baissé considérablement en utilisant notre méthode de 13.3 pixels à 1.91 pixels. Ces résultats mettent en évidence la bonne précision que notre solution assure contrairement à la méthode de [7] qui présente des imprécisions locales et globales. En effet, utiliser uniquement les points 3D associés aux modèles des bâtiments causent des imprécisions locales quand peu de bâtiments sont observables comme le montre la figure 4 où le nuage de points (en orange) n'est pas aligné avec les empreintes des bâtiments. De plus, optimiser tous les degrés de liberté à travers un ajustement de faisceaux contraint uniquement aux modèles des bâtiments résulte des imprécisions globales principalement observées au niveau des paramètres hors plan (voir figure 5).

		Erreur de re-projection (pixels)			
		Moyenne	Écart type	Max	Min
Versailles	[7]	17.85	13.30	40.90	4.11
	Nous	7.32	1.91	10.90	4.04
Saclay	[7]	14.92	7.02	30.14	5.56
	Nous	8.37	3.25	16.31	3.52

TABLE 1 – **Résultats quantitatives pour les séquences réelles.** Comparaison des erreurs de re-projection entre les coins labellisés et les coins des modèles de bâtiments re-projeté.

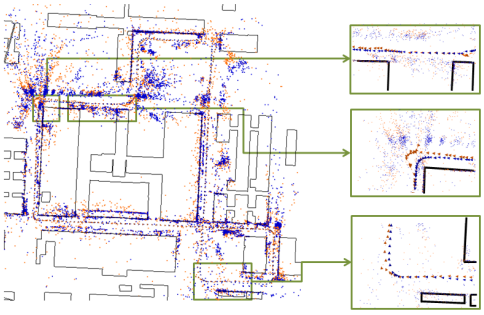
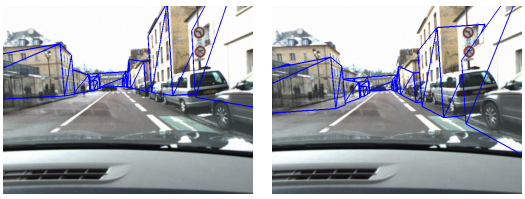


FIGURE 4 – **Comparaison avec la méthode de [7] sur la séquence Saclay.** Vue de dessus des bases de données obtenues par notre méthode (en bleu) et celles obtenues par la méthode proposée par [7] (en orange).



(a) (b)

FIGURE 5 – **Exemples de re-projection des modèles des bâtiments sur des images de la séquence de Versailles.** A gauche, le résultat obtenu par notre méthode. A droite, le résultat obtenu par la méthode [7]. Les erreurs de re-projections associées sont : (a) 6.50 pixels et (b) 13.95 pixels.

(a) (b)

FIGURE 6 – **Des exemples d'applications de RA :** projection des modèles des bâtiments, insertion des informations routières et la trajectoire du véhicule.

6 Conclusion

Dans cet article, nous avons proposé une solution précise et robuste pour modéliser automatiquement des grands environnements. Cette approche fusionne les contraintes multi-vues, celles apportées par le GPS et des modèles SIG. Notre solution est facile à déployer puisqu'elle ne nécessite pas des matériels coûteux. De plus, plusieurs bases d'amers peuvent être facilement fusionnées pour couvrir des environnements plus larges. Tous ces faits réunis rendent le déploiement de notre solution possible chez l'utilisateur final. Contrairement à la majorité des solutions existantes, notre approche permettrait ainsi une mise à jour continue et collaborative des bases d'amers. Les résultats sur les séquences réelles illustrent la précision élevée atteinte. Dans nos prochains travaux, nous nous intéresserons aux imperfections affectant l'angle de roulis qui sont parfois visibles à cause des courbures éventuelles des routes et qui ne sont pas modélisées dans le MET.

Références

[1] V. Gay-Bellile, P. Lothe, S. Bourgeois, E. Royer et S. Naudet-Collette *Augmented reality in large environments : Application to aided navigation in urban context,* ISMAR 2010.

[2] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser et P. Sayd *Real Time Localization and 3D Reconstruction,* CVPR 2006.

[3] Y. Li, N. Snavely, D. Huttenlocher et P. Fua *Worldwide pose estimation using 3d point clouds*, ECCV 2012.

[4] A. R. Zamir et M. Shah *Accurate Image Localization Based on Google Maps Street View,* ECCV 2010.

[5] R.S. Kaminsky, N. Snavely, S.M. Seitz et R. Szeliski *Alignment of 3D point clouds to overhead images,* CVPR Workshop 2009.

[6] C. Strecha, T. Pylvänäinen en P. Fua *Dynamic and scalable large scale image reconstruction,* CVPR 2010.

[7] P. Lothe, S. Bourgeois, F. Dekeyser, E. Royer et M.Dhome *Toward Large Scale Model Construction for Vision-Based Global Localisation,* VISAPP 2009.

[8] M. Lhuillier *Incremental Fusion of Structure-from-Motion and GPS Using Constrained Bundle Adjustments,* PAMI 2012.

[9] M. Tamaazousti, V. Gay-Bellile, S. Naudet-Collette, S. Bourgeois et M. Dhome *NonLinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment,* CVPR 2011.

[10] D. Larnaout, V. Gay-Bellile, S. Bourgeois et M. Dhome *Vehicle 6-DoF Localization Based on SLAM Constrained by GPS and Digital Elevation Model Information,* ICIP 2013.